Role of Donor Genital Tract HIV-1 Diversity in the Transmission Bottleneck

Peter T. Hraber, Elena E. Giorgi, T-6; Tanmoy Bhattacharya, T-2; Bette T. Korber, T-6; Debrah Boeras, Mackenzie Hurlston, Susan Allen, Cynthia Derdeyn, Eric Hunter, Emory University; Tammy Evans-Strickfaden, Clyde Hart, Centers for Disease Control and Prevention; Joseph Mulenga, Zambia Emory, HIV Research Project, Lusaka, Zambia; Etienne Karita, Projet San Francisco, Kigali, Rwanda

In the heterosexual transmission of HIV-1 infection, a genetic bottleneck is imposed on the virus quasispecies. To understand whether limited genetic diversity in the genital tract (GT) of the transmitting partner drives this bottleneck, viral sequences from blood and genital fluids of eight transmission pairs from Rwanda and Zambia were analyzed. The chronically infected transmitting partner's virus population was heterogeneous, with distinct genital subpopulations. GT populations in two of four women sampled longitudinally were stable for weeks to months. Surprisingly, the transmitted founder variant was not derived from the predominant GT subpopulations. Rather, in each case, the transmitting variant was phylogenetically distinct from the sampled locally replicating population. Although the exact distribution of the GT virus population at transmission cannot be unambiguously defined in human studies, it is unlikely that the transmission bottleneck is always driven by limited viral diversity in the donor GT or that HIV transmission is purely random.

The predominant mode of HIV-1 infection is heterosexual transmission, where a genetic bottleneck is imposed on the virus quasispecies. To understand whether limited genetic diversity in the genital tract (GT) of the transmitting partner drives this bottleneck, we analyzed viral envelope sequences from the blood and genital fluids (cervical swab or semen) of eight linked transmission pairs (both donor and recipient) from Rwanda and Zambia. The chronically infected donor's virus population was heterogeneous and predominated by distinct GT subpopulations (Fig. 1). Virus populations within the GT of two of four women sampled longitudinally exhibited stability over time intervals on the order of weeks to months. Surprisingly, the transmitted founder variant was not derived from predominant genital tract subpopulations. Rather, in each case, the transmitting variant was phylogenetically distinct from the predominant locally replicating populations in the sample. Though the exact distribution of the virus population present in the GT at the time of transmission cannot be unambiguously defined in these human studies, it is unlikely that the transmission bottleneck is driven in every case by limited viral diversity in the donor GT. Further, a quantitative test of random transmission indicates that HIV transmission is not solely a stochastic sampling from the donor GT, but more likely involves selection for some property other than abundance in the GT [1].

To address whether or not the transmitted sequence is randomly sampled from the donor virus population, we used an objective clustering criterion to relate sequences. A distance threshold D defines a cluster as any subset of sequences within distance D nucleotides from one another. That is, any two sequences occupy the same cluster if they differ at D sites or fewer. We computed pairwise distances among available GT sequences from any given donor, then assigned sequences to clusters. For any given D, and for each transmission pair i=1,…,N, we calculate the frequency f_i of donor GT sequences that do not cluster with other sequences. Let $P_D(n)$ be the probability that n donors transmit a sequence outside a cluster. Then:

$$P_D(0) = \prod_{i=1}^{N} (1 - f_i)$$

$$P_D(1) = \sum_{j=1}^{N} \prod_{i \neq j} f_j (1 - f_i) = P_D(0) \sum_{i=1}^{N} \frac{f_i}{1 - f_i}$$

and so on. In general, the probability generating function is:

$$g_D(x) = \sum_{i=1}^{N} P_D(i) x^i .$$

It follows that:

$$g_D(x) = \prod_{i=1}^{N} [(1 - f_i) + f_i x].$$

We thus calculated $P_D(n)$ using the formula:

$$P_D(n) = \frac{1}{n!} \frac{d^n}{dx^n} \left[g_D(x) \right]_{x=0}$$

For more information contact Peter T. Hraber at phraber@lanl.gov

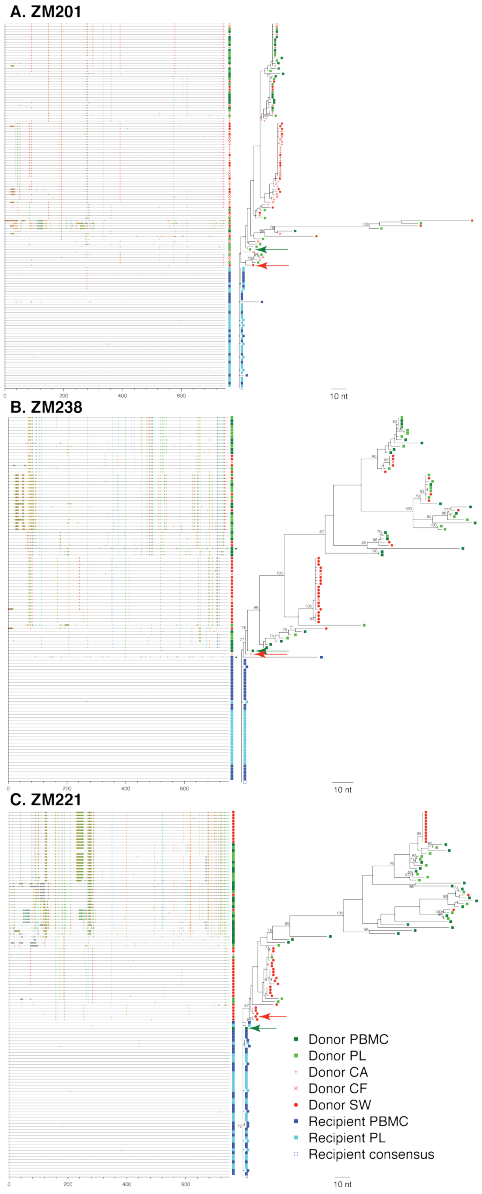

Fig. 1. Transmission analysis of HIV-1 with molecular phylogenetics shows distinct subpopulations in the donor female genital tract (GT). Though subpopulations of nearly identical sequences dominate female GT samples, the transmitted sequences are instead limited to distinct variants. For three of eight representative transmission pairs from Zambia (ZM) and Rwanda, aligned env V1-V4 nucleotide sequences are shown as phylogenetic trees (right) paired with highlighted polymorphism plots (left). Both GT (red symbols: CA, cell-associated; CF, cell-free; SW, cerival swab) and blood-borne virus (green squares: PBMC, peripheral blood mononucleocytes; PL plasma) from donors are shown. The recipient blood-borne virus (blue squares: PBMC, peripheral blood mononucleocytes; PL plasma) closely resembles sequences in the donor population, confirming epidemiological linkage. The most closely related donor sequences from blood (green arrows) and GT (red arrows) samples are indicated. Tick marks highlight locations of nucleotide differences (A, green; T, red; G, orange; C, cyan; gaps, grey) from the recipient consensus sequence (open blue square), the putative transmitted/founder virus in the recipient.

for n=1,…,8 and D=0,…,9 as above. To test for random transmission, for each D, we consider n_{obs} the observed number of donors that transmit sequences outside a cluster and compute:

$$p(D) = \sum_{n=n_{obs}}^{N} P_D(n)$$

The quantity $p(D)$ is the overall probability among eight donors that the observed number of sequences transmitted outside (not inside) a cluster is different than we would observe if transmission were to sample randomly from the donor virus population. Small values of $p(D)$ indicate that the event is unlikely to occur by chance (Fig. 2).

Software to compute the probability of non-random transmission is freely available online at ftp://ftp-t10.lanl.gov/pub/hivdb/tort/tort.tar.gz.

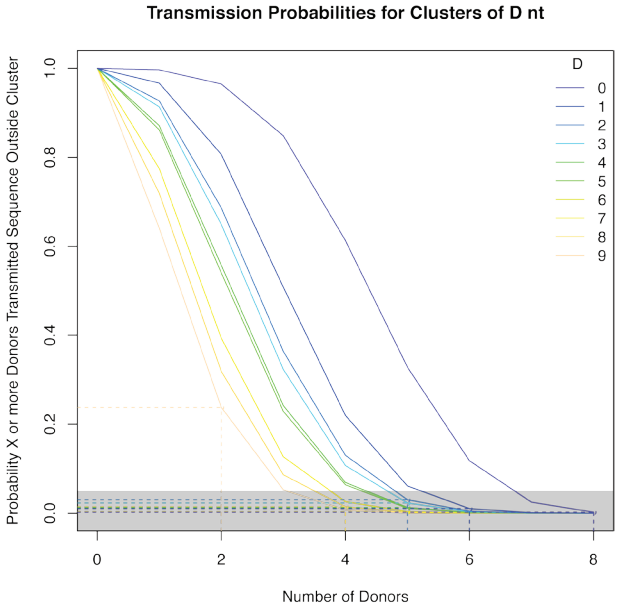

Fig. 2. Selection of transmitted/founder viruses from the donor population is not random. Subpopulations of donor GT sequences were clustered with threshold distance criterion, D, which varied from 0–9 nucleotides. For each D, the proportion of nearly identical (clustered) GT sequences was used to compute the probability that virus transmission sampled randomly from outside a cluster. Comparing these probabilities across eight transmission pairs with the number of transmission events found to occur outside a cluster (vertical dashed lines) quantifies the probability of random transmission. Small probabilities indicate non-random transmission ($p<0.05$, grey box).

[1] Boeras, D. et al., *Proc Natl Acad Sci USA* **108,** E1156 (2011).

Funding Acknowledgments
National Institute for Allergy and Infectious Disease, Center for HIV/AIDS Vaccine Immunology; NIH; LANL Laboratory Directed Research and Development Program